

A short note on a class of statistics for estimation of the Hurst index of fractional Brownian motion

K. Kubilius^{1,*,\dagger} and V. Skorniakov^{1,2,**}

¹Vilnius University, Institute of Mathematics and Informatics, Akademijos 4, LT-08663, Vilnius, Lithuania

²Vilnius University, Faculty of Mathematics and Informatics, Naugarduko 24, LT-03225, Vilnius, Lithuania

Abstract

We propose some class of statistics suitable for estimation of the Hurst index of the fractional Brownian motion based on the second order increments of an observed discrete trajectory.

Keywords: fractional Brownian motion, Hurst index, consistent estimator, central limit theorem

1 Introduction

The aim of this short note is to present some class of statistics suitable for estimation of the Hurst index from discretely observed trajectory of the fractional Brownian motion (fBm). The idea behind construction is fairly simple and makes use of self-similarity and stationarity of the second order increments. Employment of these properties immediately enables to prove usual asymptotic results, namely strong consistency and normality. Perhaps the most interesting feature of the present work is an intersection with results of [4] suggesting one of possible generalizations. It appears that in case of the fBm increment ratio (IR) statistic of [4] belongs to the class of statistics considered in the paper. Therefore it seems that following along the lines of [4] one can build a class of statistics similar to the ones considered here and suitable for measuring the roughness of random paths considered in [4].

The structure of the paper is as follows. In section 2 we state theoretical result. In section 3 we give two concrete examples from the class of suggested statistics. Finally section 4 contains a proof of the main theoretical result together with a short subsection of auxiliary results needed for the proof and collected only for the reader's convenience. The reader unfamiliar with a topic should consult that subsection first and then proceed to the main result.

2 Main result

2.1 Statement

We assume that an observed discrete sample corresponds to the uniform partition of a time interval of a trajectory $(B_t^H)_{t \in [0;T]}$ with fixed $T > 0$. Since B^H is self-similar, w.l.o.g. in the rest of the paper we concentrate on samples $B_{\frac{i}{n}}^H$, $i = 0, \dots, n$, corresponding to a trajectory $(B_t^H)_{t \in [0;1]}$. Whenever it is possible, we omit superscript and write B_t instead of B_t^H .

Let $d_{n,i} = d_i = \Delta^{(2)} B_{\frac{i+1}{n}} = B_{\frac{i+1}{n}} - 2B_{\frac{i}{n}} + B_{\frac{i-1}{n}}$, $i = 1, \dots, n-1$ be an array of the second order differences obtained from the sample and $r_i = \frac{d_{i+1}}{d_i}$, $i = 1, \dots, n-1$. Our main result is contained in the proposition given below.

PROPOSITION 2.1. *Let $\rho(x) = \frac{-7-9^x+4^{x+1}}{2(4-4^x)}$, $x \in (0;1)$, K denotes a standard Cauchy r.v.¹ and $h : \mathbb{R} \rightarrow \mathbb{R}$ be measurable. Assume that:*

- (i) $E h^2(K + \rho(H)) < \infty$;
- (ii) $H \mapsto E h(K + \rho(H))$ possesses non-zero derivative of constant sign in a neighborhood of H .

*Corresponding author. E-mail: kestutis.kubilius@mii.vu.lt

^{\dagger}This research was funded by a grant (No. MIP-048/2014) from the Research Council of Lithuania.

**E-mail: viktor.skorniakov@mif.vu.lt

¹i.e. K is absolutely continuous and has a density with respect to Lebesgue measure on $\mathcal{B}(\mathbb{R})$ given by $f_K(x) = \frac{1}{\pi(1+x^2)}$, $x \in \mathbb{R}$.

Then

$$\widehat{H}_{n,h} = \varphi(\bar{h}_n) \rightarrow H \text{ a.s.} \quad \text{and} \quad \sqrt{n}(\widehat{H}_{n,h} - H) \xrightarrow{d} N(0; \sigma_h^2), \quad \text{where} \quad \bar{h}_n = \frac{1}{n} \sum_{i=1}^n h(r_i),$$

φ denotes an inverse of $H \mapsto \mathbb{E}h(K + \rho(H))$ and σ_h^2 is precisely defined in subsection 4.2.

3 Concrete examples

In this section we give two examples of functions which satisfy conditions (i) – (ii) stated in proposition 2.1. The first function is considered because it usually happens that arcsin transform symmetrizes distribution and improves normal approximation. The second one demonstrates a connection with [4] discussed in the introduction. It is worthwhile to mention that in [4] the reader can find an example of applications for estimation problems within a framework of diffusions. Statistics introduced in this paper may be applied in a similar way.

In case of the first example we check conditions (i) – (ii) and give an expression for φ . In case of the second example we simply state a form of h and $\mathbb{E}h(K + \rho(H))$ referring for the details to [4].

Before proceeding to the mentioned examples note that $x \mapsto \rho(x)$ is increasing with a range equal to $(-\frac{2}{3}; -2 + \frac{9}{8} \frac{\ln 9}{\ln 4})$ and derivative

$$\rho'(x) = \frac{1}{2} \left(\frac{-9^x \cdot 4 \cdot \ln(9) + 4^x \cdot 9 \cdot \ln(4) + 36^x \ln(9/4)}{(4 - 4^x)^2} \right). \quad (3.1)$$

In the rest of this subsection we omit an argument for $\rho(H)$ when it appears unnecessary and write ρ instead.

Example 1. Let $h(x) = \sin x$. Then h is bounded and (i) holds. Next, note that K is symmetric r.v. Hence, its characteristic function $\psi_K(t) = \mathbb{E} \cos(tK) = e^{-|t|}$ and for any odd g it holds that $\mathbb{E}g(K) = 0$. Therefore

$$\mathbb{E} \sin(K + \rho) = \cos \rho \mathbb{E} \sin K + \sin \rho \mathbb{E} \cos K = (\sin \rho) \psi_K(1) = \frac{\sin \rho}{e}.$$

Properties of $H \mapsto \rho(H)$ imply that $H \mapsto e^{-1} \sin \rho(H)$ is increasing on $(0; 1)$ with an inverse

$$\varphi(y) = \rho^{-1}(\arcsin(ey)), \quad y \in \left(\frac{\sin(-2/3)}{e}; \frac{\sin(-2 + \frac{9}{8} \frac{\ln 9}{\ln 4})}{e} \right)$$

and derivative $\varphi'(\mathbb{E} \sin(K + \rho(H))) = e(\rho'(H) \cos \rho(H))^{-1}$.

Example 2. Setting $h(x) = \frac{1+x}{1+|x|}$ one gets statistic of [4] with

$$\mathbb{E}h(K + \rho) = \frac{1}{\pi} \left(\arccos(-\rho) + \sqrt{\frac{1+\rho}{1-\rho}} \ln \left(\frac{2}{1+\rho} \right) \right).$$

4 Auxiliary facts and the proof

4.1 Auxiliary facts

Below we list several facts needed for the proof of the main result.

4.1.1 Properties of the fBm

- Fractional Brownian motion $(B_t^H)_{t \geq 0}$ is a centered continuous Gaussian process with a covariance function

$$\mathbb{E} B_t^H B_s^H = \frac{1}{2} \left(|t|^{2H} + |s|^{2H} - |t-s|^{2H} \right), \quad H \in (0; 1).$$

- A sequence of the second order increments $X_i = \Delta^{(2)} B_{i+1}^H = B_{i+1}^H - 2B_i^H + B_{i-1}^H$, $i \geq 1$ is stationary, $\forall i$ $X_i \sim N(0; 4 - 4^H)$, $\text{Corr}(X_i, X_{i+1}) = \rho(H) = \frac{-7-9^H+4^{H+1}}{2(4-4^H)}$. Moreover (see, e.g. [3]),

$$\rho_k = \rho_k(H) = \text{Corr}(X_i, X_{i+k}) = O\left(\frac{1}{k^{4-2H}}\right), \quad k \rightarrow \infty. \quad (4.1)$$

- fBm is self-similar, that is $(B_{at}^H)_{t \geq 0} \stackrel{d}{=} (a^H B_t^H)_{t \geq 0}$.

4.1.2 Central limit theorem for a stationary sequence of Gaussian random vectors

Let $(Z_i)_{i \geq 1}$, $Z_i = (Z_{i,1}, \dots, Z_{i,d})^T$ be a stationary sequence of centered \mathbb{R}^d valued Gaussian r.v.s. For $k \in \mathbb{Z}$ and $p, q \in \{1, \dots, d\}$ denote

$$r^{(p,q)}(k) = \mathbb{E} Z_{m,p} Z_{m+k,q},$$

where m is any natural number satisfying $m, m+k \geq 1$. Note that

$$r^{(p,q)}(0) = \mathbb{E} Z_{m,p} Z_{m,q} = \mathbb{E} Z_{1,p} Z_{1,q} \quad (4.2)$$

and

$$r^{(p,q)}(-k) = \mathbb{E} Z_{m,p} Z_{m-k,q} = \mathbb{E} Z_{k+1,p} Z_{(k+1)-k,q} = \mathbb{E} Z_{k+1,p} Z_{1,q} = r^{(q,p)}(k), \quad \text{for } k \geq 1. \quad (4.3)$$

We make use of the following result given in [1] (Theorem 2).

THEOREM 4.1. *Assume that $f : \mathbb{R}^d \rightarrow \mathbb{R}$ is measurable, $\mathbb{E} f^2(Z_1) < \infty$ and for each $(p, q) \in \{(i, j) \mid i, j = 1, \dots, d\}$ there exist finite limits*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j,k=1}^n r^{(p,q)}(j-k), \quad \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j,k=1}^n (r^{(p,q)}(j-k))^2.$$

Then

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n (f(Z_i) - \mathbb{E} f(Z_1)) \xrightarrow{d} N(0; \sigma_f^2), \quad (4.4)$$

where

$$\sigma_f^2 = \text{Var}(f(Z_1)) + 2 \sum_{k=1}^{\infty} \text{cov}(f(Z_1), f(Z_{1+k})).$$

4.2 Proof

Retain the notions introduced in the previous sections and consider a bivariate Gaussian sequence $Z_i = (Z_{i,1}, Z_{i,2}) = \frac{1}{\sqrt{4-4H}}(X_{i+1}, X_i)$, $i \geq 1$. It follows from the properties listed above that $(Z_i)_{i \geq 1}$ is stationary and that

$$|r^{(p,q)}(k)| = |\mathbb{E} Z_{1,p} Z_{1+k,q}| \leq \frac{C}{k^{4-2H}} \quad (4.5)$$

for all $p, q = 1, 2$, $k \geq 1$, and finite positive constant C depending only on H . The bound implies convergence of $\sum_{k=1}^{\infty} r^{(p,q)}(k)$ and $\sum_{k=1}^{\infty} (r^{(p,q)}(k))^2$. By (4.2)–(4.3)

$$\frac{1}{n} \sum_{j,k=1}^n (r^{(p,q)}(j-k))^i = (r^{(p,q)}(0))^i + \frac{1}{n} \left(\sum_{j=1}^{n-1} \sum_{k=j+1}^n (r^{(q,p)}(k-j))^i + \sum_{j=1}^{n-1} \sum_{k=j+1}^n (r^{(p,q)}(k-j))^i \right), \quad i = 1, 2.$$

Thus, it suffices to show that for any $p, q = 1, 2$, there exist limits

$$\lim_{n \rightarrow \infty} n^{-1} \sum_{j=1}^{n-1} \sum_{k=j+1}^n (r^{(p,q)}(k-j))^i, \quad i = 1, 2.$$

Let $i = 1$. Fix p, q and note that

$$\begin{aligned} \frac{1}{n} \sum_{j=1}^{n-1} \sum_{k=j+1}^n r^{(p,q)}(k-j) &= [k-j=l] = \frac{1}{n} \sum_{j=1}^{n-1} \sum_{l=1}^{n-j} r^{(p,q)}(l) = \sum_{l=1}^{n-1} r^{(p,q)}(l) \left(1 - \frac{l}{n}\right) \\ &= \left[\mathbf{1}_{\{1, \dots, n-1\}}(l) r^{(p,q)}(l) \left(1 - \frac{l}{n}\right) = \psi_n(l) \right] = \sum_{l=1}^{\infty} \psi_n(l) = \int_{\mathbb{N}} \psi_n d\mu, \end{aligned}$$

where μ denotes a counting measure on \mathbb{N} . For each fixed l it holds true $\psi_n(l) \xrightarrow[n \rightarrow \infty]{} r^{(p,q)}(l)$. By (4.5),

$$|\psi_n(l)| \leq \psi(l) = \frac{C}{l^{4-2H}}, \quad l \in \mathbb{N}.$$

Since ψ is integrable with respect to μ , Dominated Convergence theorem yields relationship

$$\lim_{n \rightarrow \infty} \int_{\mathbb{N}} \psi_n d\mu = \int_{\mathbb{N}} \lim_{n \rightarrow \infty} \psi_n d\mu = \sum_{l=1}^{\infty} r^{(p,q)}(l) \in \mathbb{R}.$$

Hence,

$$\lim_{n \rightarrow \infty} n^{-1} \sum_{j=1}^{n-1} \sum_{k=j+1}^n r^{(p,q)}(k-j) = \sum_{l=1}^{\infty} r^{(p,q)}(l).$$

Identical argument shows that

$$\lim_{n \rightarrow \infty} n^{-1} \sum_{j=1}^{n-1} \sum_{k=j+1}^n (r^{(p,q)}(k-j))^2 = \sum_{l=1}^{\infty} (r^{(p,q)}(l))^2.$$

Therefore theorem 4.1 applies to (Z_i) provided function f is suitably chosen. Let

$$f(x, y) = \mathbf{1}_{\mathbb{R} \times (\mathbb{R} \setminus \{0\})}(x, y) h\left(\frac{x}{y}\right), \quad R_i = \frac{Z_{i,1}}{Z_{i,2}}, \quad i \geq 1.$$

Recall that a ratio of two independent standard Gaussian r.v.s. has the standard Cauchy distribution. Keeping this in a view we arrive to the following conclusions² :

- $R_i = \frac{Z_{i,1}}{Z_{i,2}} = \frac{(X_{i+1} - \rho(H)X_i)/\sqrt{4-4^H}}{X_i/\sqrt{4-4^H}} + \rho(H) \stackrel{d}{=} K + \rho$, since $(X_{i+1} - \rho(H)X_i)/\sqrt{4-4^H}$, $X_i/\sqrt{4-4^H} \sim N(0; 1)$ are uncorrelated;
- by self-similarity $(R_i) \stackrel{d}{=} (r_i)$;
- because of absolute continuity of Gaussian r.v. and the previous facts $E f^j(Z_1) = E h^j(R_1) = E h^j(r_1) = E h^j(K + \rho(H))$, $j = 1, 2$.

Also note that in addition to stationarity $(X_i)_{i \geq 1}$ has vanishing correlation (see (4.1)). Therefore it is ergodic and conclusions listed above together with Ergodic theorem (see Corollary 8.6.3 in [2]) yield relationships

$$1 = P\left(\lim_{n \rightarrow \infty} \frac{f(Z_1) + \dots + f(Z_n)}{n} = E f(Z_1)\right) = P\left(\lim_{n \rightarrow \infty} \frac{h(r_1) + \dots + h(r_n)}{n} = E h(r_1) = E h(R_1)\right).$$

That is, $\bar{h}_n \rightarrow E h(R_1)$ a.s. and by continuous mapping theorem $\varphi(\bar{h}_n) \rightarrow \varphi(E h(R_1)) = \varphi(E h(K + \rho(H))) = H$ a.s.

Next, note that from previously stated equality $E f^2(Z_1) = E h^2(K + \rho(H))$ and condition $E h^2(K + \rho(H)) < \infty$ it follows that Theorem 4.1 applies to the chosen f and gives (4.4) which may be rewritten as

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_{i=1}^n (h(r_i) - E h(r_1)) &= \sqrt{n}(\bar{h}_n - E h(r_1)) \xrightarrow{d} N(0; \sigma_f^2), \\ \sigma_f^2 &= \text{Var}(h(R_1)) + 2 \sum_{k=1}^{\infty} \text{cov}(h(R_1), h(R_{1+k})). \end{aligned} \tag{4.6}$$

To finish the proof, one has to apply the Delta method, which also yields asymptotic variance $\sigma_h^2 = (\varphi'(E h(K + \rho(H))))^2 \sigma_f^2$. \square

References

- [1] M. A. Arcones, Limit theorems for nonlinear functionals of a stationary Gaussian sequence of vectors, *The Annals of Probability*, **22**(4) (1994), 2242-2274.
- [2] Athreya, Krishna B. and Lahiri, Soumen N., *Measure Theory and Probability Theory* (Springer Texts in Statistics), 2006, Springer-Verlag New York, Inc.
- [3] J.-F. Coeurjolly, Estimating the parameters of a fractional Brownian motion by discrete variations of its sample paths, *Statistical Inference for Stochastic Processes*, **4** (2001), 199-227.
- [4] J. M. Bardet and D. Surgailis, Measuring the roughness of random paths by increment ratios, *Bernoulli*, **17**(2) (2011), 749-780.

² K denotes a r.v. having standard Cauchy distribution (see subsection 2)